AMENDMENTS TO THE CLAIMS

Please amend the Claims as follows:

1. (Original) A computer-based method ~~for~~ to perform query optimization by automatically finding and exploiting hidden, fuzzy algebraic constraints in a database, said method comprising the steps of:

(a) constructing one or more candidates of form $C=(a_1, a_2, P, \oplus)$, wherein a_1 and a_2 are numerical attributes associated with column values of data in said database, P is a pairing rule, and $\oplus$ is any of the following algebraic operators: +, -, ×, or /;

(b) constructing, for each candidate identified in (a), an algebraic constraint $AC=(a_1, a_2, P, \oplus, I_1,\ldots, I_k)$ by applying ~~any of, or a combination of the following techniques to a sample of column values: statistical histogramming,~~ a segmentation~~, or clustering~~ technique, where $I_1, \ldots, I_k$ is a set of disjoint intervals and $k \geq 1$, said step of constructing algebraic constraint further comprising the steps of:

constructing a sample set W_C of an induced set Ω_C, wherein P is a join predicate between tables R and S and $\Omega_C = \{r.a_1 \oplus r.a_2 : r \in R\}$ when the pairing rule P is a trivial rule $\varnothing_R$ and $\Omega_C = \{r.a_1 \oplus s.a_2 : r \in R, s \in S, \text{and } (r,s) \text{ satisfies } P\}$;

sorting n data points in said sampled set W_C in increasing order as $x_1 \leq x_2 \leq \ldots \leq x_n$ and constructing a set of disjoint intervals $I_1, \ldots, I_k$ such that data in sample W_C falls within one of said disjoint intervals, wherein segmentation for constructing said set of disjoint intervals is specified via a vector of indices (*i(1)*,

$i(2)$, ..., $i(k)$) and the j^{th} interval is given by $I_j=[x_{i(j-1)+1}, x_{i(j)}]$ and length of Ij, denoted by Lj, is given by $L_j = x_{i(J)} - x_{i(J-1)+1}$; and

wherein the function for optimizing cost associated with said segmentation is $c(S) = wk + (1-w)\left[\frac{1}{\Delta}\sum_{j=1}^{k} L_j\right]$ with w being a fixed weight between 0 and 1 and a segmentation that minimizes c is defined by placing adjacent points x_I and x_{I+1} in the same segment if and only if $x_{I+1}-x_I < d^*$, where $d^* = \Delta(w/(1-w))$, and

wherein said constructed algebraic constraints are used in query optimization.

2. (Original) A compute-based method as per claim 1, wherein one or more pruning rules are used to limit said number of constructed candidates.

3. (Original) A computer-based method as per claim 2, wherein said pairing rule P represents either a trivial pairing rule $\emptyset_R$ or a join between tables R and S and said pruning rules comprise any of, or a combination of the following:

pairing rule P is of form $R.a = S.b$ or of the form $\emptyset_R$, and the number of rows in either table R or table S lies below a specified threshold value;

pairing rule P is of form $R.a = S.b$ with $a \in K$ and the number of distinct values in $S.b$ divided by the number of values in $R.a$ lies below a specified threshold value, wherein K is a set comprising key-like columns among all columns in said database;

pairing rule P is of form $R.a = S.b$, and one or both of R and S fails to have an index on any of its columns; or

pairing rule P is of form $R.a = S.b$ with $a \in K$, and $S.b$ is a system-generated key.

4. (Original) A computer-based method as per claim 1, wherein said method further comprises the steps of:

identifying a set of useful algebraic constraints via one or more pruning rules; and

partitioning data into compliant data and exception data.

5. (Original) A computer-based method as per claim 4, wherein said method further comprises the steps of:

receiving a query;

modifying said query to incorporate identified constraints; and

combining results of modified query executed on data in said database and said original query executed on exception data.

6. (Original) A computer-based method as per claim 4, wherein said partitioning is done by incrementally maintained materialized views, partial indices, or physical partitioning of the table.

7. (Original) A computer-based method as per claim 2, wherein said pruning rules comprise any of, or a combination of the following:

a_1 and a_2 are not comparable data types;

the fraction of NULL values in either a_1 or a_2 exceeds a specified threshold; or

either column a_1 or a_2 is not indexed.

8. (Original) A computer-based method as per claim 1, wherein said step of constructing one or more candidates further comprises the steps of:

generating a set P of pairing rules; and

for each pairing rule $P \in \mathbb{P}$, systematically considering possible attribute pairs (a_1, a_2) and operators $\oplus$ with which to construct candidates.

9. (Original) A computer-based method as per claim 8, wherein said step of generating a set $\mathbb{P}$ of pairing rules further comprises the steps of:

initializing $\mathbb{P}$ to be an empty set;

adding a trivial pairing rule of the form $\varnothing_R$ to said set $\mathbb{P}$ for each table R in said database; and

generating and adding nontrivial pairing rules to said set $\mathbb{P}$ based upon identifying matching columns via an inclusion dependency, wherein a column b is considered a match for column a if:

data in columns a and b are of a comparable type; or

either (i) column a is a declared primary key and column b is a declared foreign key for the primary key, or (ii) every data value in a sample from column b has a matching value in column a.

10. (Original) A computer-based method as per claim 8, wherein said step of generating a set $\mathbb{P}$ of pairing rules further comprises the steps of:

initializing $\mathbb{P}$ to be an empty set;

adding a trivial pairing rule of the form $\varnothing_R$ to said set $\mathbb{P}$ for each table R in said database; and

generating a set K of key-like columns from among all columns in said database with each column in set K belonging to a predefined set of types T, said set K comprising declared

primary key columns, declared unique key columns, and undeclared key columns, wherein said primary keys or declared unique keys are compound keys of form $a = (a_1, ..., a_m) \in T^m$ for $m>1$;

adding nontrivial pairing rules to said set P based upon identifying matching compound columns via an inclusion dependency wherein, given a compound key $(a_1, ..., a_m) \in K$, a compound column b is considered a component wise match for compound column a if:

data in compound columns a and b are of a comparable type; or

either (i) compound column a is a declared primary key and compound column b is a declared foreign key for the primary key, or (ii) every data value in a sample from compound column b has a matching value in compound column a.

11. (Cancelled)

12. (Currently Amended) A computer-based method as per ~~claim 11~~claim 1, wherein widths associated with said intervals are expanded to avoid additional sampling required to increase right end point to equal maximum value in Ω_C.

13. (Currently Amended) A computer-based method as per ~~claim 11~~claim 1, wherein size of said sampled set is approximated via the following iterative steps:

(a) given a k-segmentation, setting counters i=1 and k=1;

(b) selecting a sample size $n=n^*$, wherein $n^*(k) \approx \frac{\chi^2_{1-p}(2-f)}{4f} + \frac{k}{2}$, wherein p is the probability that at least a fraction of points in Ω_C that lie outside the intervals is at most f;

(c) obtaining a sample based on (b), computing algebraic constraints, and identifying a number k' of bump intervals; and

(d) if $n \geq n^*(k')$ or $i = i_{max}$, then utilizing sample size in (b); else setting counters $k=k'$ and $i=i+1$, and returning to step (b).

14. (Cancelled).

15. (Cancelled).

16. (Original) A computer-based method as per claim 1, wherein said method is implemented across networks.

17. (Original) A computer-based method as per claim 16, wherein said across networks element comprises any of, or a combination of the following: local area network (LAN), wide area network (WAN), or the Internet.

18. (Cancelled).

19. (Cancelled).

20. (Cancelled).

21. (Cancelled).

22. (Currently Amended) An article of manufacture comprising a computer usable medium having computer readable program code embodied therein which implements a method to perform query optimization by ~~for~~ automatically finding and exploiting hidden, fuzzy algebraic constraints in a database, said method comprising the steps of:

(a) computer readable program code constructing one or more candidates of form $C=(a_1, a_2, P, \oplus)$, wherein a_1 and a_2 are numerical attributes associated with column values of data in said database, P is a pairing rule, and $\oplus$ is any of the following algebraic operators: +, -, ×, or /;

(b) computer readable program code constructing, for each candidate identified in (a), an algebraic constraint $AC=(a_1, a_2, P, \oplus, I_1,\ldots, I_k)$ by applying ~~any of, or a combination of the following techniques to a sample of column values: statistical histogramming,~~ a segmentation technique~~, or clustering~~, where $I_1, \ldots, I_k$ is a set of disjoint intervals and $k \geq 1$, said step of constructing algebraic constraint further comprising the steps of:

constructing a sample set W_C of an induced set Ω_C, wherein P is a join predicate between tables R and S and $\Omega_C = \{r.a_1 \oplus r.a_2 : r \in R\}$ when the pairing rule P is a trivial rule $\varnothing_R$ and $\Omega_C = \{r.a_1 \oplus s.a_2 : r \in R, s \in S, and\ (r,s)\ satisfies\ P\}$;

sorting n data points in said sampled set W_C in increasing order as $x_1 \leq x_2 \leq \ldots \leq x_n$ and constructing a set of disjoint intervals $I_1, \ldots, I_k$ such that data in sample W_C falls within one of said disjoint intervals, wherein segmentation for constructing said set of disjoint intervals is specified via a vector of indices $(i(1), i(2), \ldots, i(k))$ and the j^{th} interval is given by $I_j=[x_{i(j-1)+1}, x_{i(j)}]$ and length of Ij, denoted by Lj, is given by $L_j = x_{i(j)} - x_{i(j-1)+1}$; and

wherein the function for optimizing cost associated with said segmentation is $c(S) = wk + (1-w)\left[\frac{1}{\Delta}\sum_{j=1}^{k} L_j\right]$ with w being a fixed weight between 0 and 1 and a segmentation

that minimizes c is defined by placing adjacent points x_l and x_{l+1} in the same segment if and only if $x_{l+1}-x_l < d^*$, where $d^* = \Delta(w/(1-w))$, and

wherein said constructed algebraic constraints are used in query optimization.

23. (Original) An article of manufacture as per claim 22, wherein said medium further comprises:

computer readable program code identifying a set of useful algebraic constraints via heuristics comprising a set of pruning rules; and

computer readable program code partitioning data into compliant data and exception data.

24. (Original) An article of manufacture as per claim 23, wherein said medium further comprises:

computer readable program code aiding in receiving a query;

computer readable program code modifying said query to incorporate identified constraints; and

computer readable program code combining results of modified query executed on data in said database and said original query executed on exception data.

Please cancel claims **25-38**.